ELSEVIER

Theoretical Computer Science 218 (1999) 83–94

Theoretical Computer Science

Fine and Wilf's theorem for three periods and a generalization of Sturmian words

M.Gabriella Castelli *, Filippo Mignosi, Antonio Restivo

Dipartimento di Matematica ed Applicazioni, Università di Palermo via Archirafi, 34, 90123 Palermo, Italy

Abstract

We extend the theorem of Fine and Wilf to words having three periods. We then define the set *3-PER* of words of maximal length for which such result does not apply. We prove that the set *3-PER* and the sequences of complexity $2n+1$, introduced by Arnoux and Rauzy to generalize Sturmian words, have the same set of factors. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Combinatorics on words; Periodicity; Euclid's algorithm; Sturmian words

1. Introduction

The notion of periodicity plays an important role in combinatorics on words. A central result in the investigation on periodicity is the Fine and Wilf's theorem. This theorem states that, if a word w has periods p, q and its length $|w|$ is greater than or equal to $p+q-\gcd(p,q)$, then w has also period $\gcd(p,q)$.

With reference to this result, de Luca and Mignosi introduced in [6] the set *PER* of all words having periods p, q, which are coprimes and such that $|w| = p+q-2$. Thus a word w belongs to *PER* if it is a power of a single letter or is a word of maximal length for which the theorem of Fine and Wilf does not apply. In [6] de Luca and Mignosi, in particular, proved that the set of factors of *PER* coincides with the set of finite factors of Sturmian words. This result is at origin of a deep investigation on the combinatorial properties of Sturmian words carried on in several papers (cf. [3, 7, 8, 18]).

In this paper we first extend the theorem of Fine and Wilf to words having three periods. We introduce a function $f(x,y,z)$ and we prove that, if a word w has periods p_1, p_2, p_3 and length $|w| \geqslant f(p_1,p_2,p_3)$, then w has also period $\gcd(p_1,p_2,p_3)$. This result includes, as a particular case, the classical Fine and Wilf's theorem, since the function $f(x,y,z)$ is such that $f(0,p,q) = p+q-\gcd(p,q)$.

* Corresponding author. E-mail: {gabry,mignosi,restivo}@altair.math.unipa.it.

PII: S0304-3975(98)00251-5

We then introduce the set *3-PER* of words w having periods p_1, p_2, p_3, which are coprimes and satisfy a supplementary condition, and such that $|w| = f(p_1, p_2, p_3) - 1$. Thus the elements of *3-PER* are the words of maximal length for which previous theorem does not apply. We will prove that the elements of *3-PER* are ternary words, i.e., are words over an alphabet of three letters.

A generalization of Sturmian words to alphabet with three letters is possible in different ways and several formal definitions have been proposed in the literature (cf. [1, 2, 13, 14, 19]).

In [1] Arnoux and Rauzy introduced sequences, on an alphabet of three letters, having complexity $2n + 1$, as generalization of Sturmian sequences. They gave a geometric representation of such sequences, generalizing a classical result on representation of Sturmian sequences by rotation.

We prove that the set of factors of *3-PER* coincides with the set of finite factors of the Arnoux and Rauzy sequences. This result, together with the result of [1], indicates the *robustness* of the notion introduced by Arnoux and Rauzy and that it can be taken as the natural candidate for being the "good" generalization of Sturmian words to three letters alphabets.

2. Preliminaries

Let us recall some basic definitions and results.

Let $w = a_1 \ldots a_n$ be a word over an alphabet $\mathscr{A}$. Denote by $|w| = n$ the length of the word w. A *period* of w is a positive integer p such that

$$a_i = a_{i+p}, \quad i \in [1, n-p].$$

The smallest p satisfying previous equalities is called *the* period of w and it is denoted by $p(n)$. Remark that any $q \geqslant |w|$ is considered a period of w. A *factor* of a word is any consecutive sequence of its symbols. If L is a set of words, denote by $F(L)$ the set of factors of words in L. Denote by $pref_k(w)$ ($suff_k(w)$, resp.) the prefix (the suffix, resp.) of length k of the word w. It is easy to verify that a word w of length n has period p if and only if $pref_{n-p}(w) = suff_{n-p}(w)$. The next two Lemmas will be useful in the sequel.

Lemma 2.1. *Let p and q be two positive integers such that $p < q$. Let $w = w_1 \ldots w_n \in \mathscr{A}^*$ be a word having periods p and q. Then the prefix (and the suffix) of w of length $n - p$ has period $q - p$ (and p, obviously).*

Proof. In the case $n - p \geqslant q - p$, (otherwise it is trivial) we have to show that

$$pref_{+(n-p)-(q-p)}(w) = suff_{(n-p)-(q-p)}(w).$$

Let $v = pref_{n-p}(w)$. Since w has period q and any prefix of v is also a prefix of w, the prefix having length $(n - p) - (q - p) = n - q$ of v is equal to the suffix of w

having the same length. Since w has period p, any suffix of w having length smaller than or equal to $|v|$ is also a suffix of v. □

Lemma 2.2. *Let p and q be two positive integers. Let $u = u_1 \dots u_n \in \mathscr{A}^*$ be a word having periods p and q. Then the word of length $n + p$*

$$w = pref_p(u)u$$

has periods p and $p + q$.

Proof. By construction the word w has period p, since the word u has period p. Let $p + q = p'$. In the case $|w| > p'$, (otherwise it is trivial) we have to prove that

$$pref_{|w|-p'}(w) = suff_{|w|-p'}(w).$$

Let us observe that, by construction, any suffix of w of length less than or equal to $|u|$ is also suffix of u and any prefix of u is also prefix of w because w has period p. We have that

$$suff_{|w|-p'}(w) = suff_{|u|-(p'-p)}(u) = pref_{|u|-(p'-p)}(u) = pref_{|w|-p'}(w),$$

because u has period $p' - p = q$ and $|w| - p' = |u| + p - p' = |u| - (p' - p) < |u|$. □

Let $\sigma : \mathbb{N} \to \mathscr{A}$ be an infinite word. Denote by $L(\sigma)$ the set of finite factors of the word σ. The *complexity* of the infinite word σ is the function

$$P_\sigma(n) = Card(L(\sigma) \cap \mathscr{A}^n).$$

A finite word w is a *right special factor* of an infinite word σ if

$$Card\{wa \in L(\sigma) | a \in \mathscr{A}\} > 1.$$

The number $Card\{wa \in L(\sigma) | a \in \mathscr{A}\}$ is called the *right degree* of w. In a similar way, we define left special factors and left degrees. For all undefined notions and notations cf. [17].

3. Euclid's algorithm for three integers

In this section we shortly report the usual algorithm to compute the gcd of three integers ([5, 16, 20]).

Let $\underline{p} = (p_1, p_2, p_3)$ be a triple of nonnegative integers. If $p_1 \leqslant p_2 \leqslant p_3$, we call $\underline{p}$ an *ordered triple*.

We consider the following operators:

(i) the operator R is defined on ordered triples as follows:

$$R(\underline{p}) = R(p_1, p_2, p_3) = \begin{cases} (p_1, p_2 - p_1, p_3 - p_1) & \text{if } p_1 \neq 0, \\ (0, p_2, p_3 - p_2) & \text{if } p_1 = 0. \end{cases}$$

(ii) the operator O, acting on an arbitrary triple, gives the corresponding ordered triple;
(iii) the operator S is then defined as

$$S(\underline{p}) = O(R(\underline{p})).$$

Given an ordered triple $\underline{p}$, let us consider the sequence $(\underline{p}^{(k)})_{k\geqslant 0}$ of ordered triples defined recursively as follows:

$$\underline{p}^{(0)} = \underline{p},$$
$$\underline{p}^{(k+1)} = S(\underline{p}^{(k)}), \quad k\geqslant 0.$$

The elements of the triple $\underline{p}^{(k)}$ are denoted by

$$\underline{p}^{(k)} = (p_1^{(k)}, p_2^{(k)}, p_3^{(k)}).$$

Let us denote by $|\underline{p}|$ the sum of the elements of the triple $\underline{p}$, i.e.,

$$|\underline{p}| = p_1 + p_2 + p_3.$$

Let

$$m(\underline{p}) = \min\{k \mid p_1^{(k)} = 0\},$$
$$M(\underline{p}) = \min\{k \mid p_1^{(k)} = p_2^{(k)} = 0\}.$$

With these notations, given the triple $\underline{p} = (p_1, p_2, p_3)$,

$$\gcd(p_1, p_2, p_3) = |\underline{p}^{(M(\underline{p}))}|.$$

We also introduce the following function, which plays an important role in this paper:

$$h(p_1, p_2, p_3) = |\underline{p}^{(m(\underline{p}))}|.$$

By definition, the function h satisfies the condition:

$$h(p_1, p_2, p_3) = h(p_1, p_2 - p_1, p_3 - p_1).$$

Definition 3.1. A triple $\underline{p} = (p_1, p_2, p_3)$ is called an homogeneous triple if

$$h(p_1, p_2, p_3) = \gcd(p_1, p_2, p_3),$$

i.e., if $m(\underline{p}) = M(\underline{p})$.

Example 3.1. Consider the triple $\underline{p} = (7, 11, 13)$. The Euclid's algorithm gives the following sequence of triples:

$$\underline{p}^{(0)} = (7, 11, 13),$$
$$\underline{p}^{(1)} = (4, 6, 7),$$
$$\underline{p}^{(2)} = (2, 3, 4),$$

$$\underline{p}^{(3)} = (1,2,2),$$
$$\underline{p}^{(4)} = (1,1,1),$$
$$\underline{p}^{(5)} = (0,0,1).$$

We have that $\gcd(7,11,13) = h(7,11,13) = 1$, i.e., $(7,11,13)$ is an homogeneous triple.

4. A periodicity theorem for words with three periods

Let us recall the classical Fine and Wilf theorem (cf. [9, 17]).

Theorem 4.1. *Let w be a word over the alphabet $\mathcal{A}$, having periods p_1 and p_2. If $n \geqslant p_1 + p_2 - \gcd(p_1, p_2)$ then the word w has also period $\gcd(p_1, p_2)$.*

In this section we consider an analogous result for words having three periods.

Given an ordered triple $\underline{p} = (p_1, p_2, p_3)$ of nonnegative integers, we introduce the function

$$f(p_1,p_2,p_3) = \tfrac{1}{2}[p_1 + p_2 + p_3 - 2\gcd(p_1,p_2,p_3) + h(p_1,p_2,p_3)].$$

Theorem 4.2. *Let w be a word over the alphabet $\mathcal{A}$ having three periods p_1, p_2 and p_3, with $p_1 \leqslant p_2 \leqslant p_3$. If $|w| \geqslant f(p_1,p_2,p_3)$, then w has also period $\gcd(p_1,p_2,p_3)$.*

Remark 4.1. The statement of this theorem includes, as a particular case, the statement of classical Fine and Wilf theorem. Indeed, the condition that a word w has periods p,q, with $p \leqslant q$, corresponds to the triple $(0,p,q)$. Since, by definition, $h(0,p,q) = p+q$, it follows that

$$f(0,p,q) = \tfrac{1}{2}[p+q-2\gcd(p,q)+p+q] = p+q-\gcd(p,q).$$

Proof of Theorem 4.2. We shall prove the theorem by induction on the integer

$$n = p_1(p_1 + p_2 + p_3).$$

The case $n = 0$ corresponds to the classical Fine and Wilf theorem (cf. Remark 4.1). Let us now suppose that the statement is true for all ordered triples $\underline{q} = (q_1,q_2,q_3)$ such that $m = q_1(q_1+q_2+q_3) < n$ and consider an ordered triple $\underline{p} = (p_1,p_2,p_3)$ such that $p_1(p_1+p_2+p_3) = n$.

Let w be a word having periods p_1, p_2 and p_3 and length $|w| \geqslant f(p_1,p_2,p_3)$.

Let u be the prefix of w of length p_1:

$$w = uv, \quad \text{with } |u| = p_1, \quad |v| = |w| - p_1.$$

By Lemma 2.1, the word v has periods p_1, $p_2 - p_1$, $p_3 - p_1$ and length $|v| = |w| - p_1$. Since $|w| \geqslant f(p_1, p_2, p_3)$, one has:

$$\begin{aligned}
|v| &= |w| - p_1 \geqslant f(p_1, p_2, p_3) - p_1 \\
&= \tfrac{1}{2}[p_1 + p_2 + p_3 - 2\gcd(p_1, p_2, p_3) + h(p_1, p_2, p_3)] - p_1 \\
&= \tfrac{1}{2}[p_1 + (p_2 - p_1) + (p_3 - p_1) - 2\gcd(p_1, p_2 - p_1, p_3 - p_1) \\
&\qquad + h(p_1, p_2 - p_1, p_3 - p_1)] \\
&= f(p_1, p_2 - p_1, p_3 - p_1).
\end{aligned}$$

By the inductive hypothesis v also has period

$$\gcd(p_1, p_2 - p_1, p_3 - p_1) = \gcd(p_1, p_2, p_3).$$

By Lemma 2.2, the word $w = uv$ has period $\gcd(p_1, p_2, p_3)$. This concludes the proof. □

Remark 4.2. One can verify that the function f satisfies the following condition: If $p_1 \leqslant p_2 \leqslant p_3$ and $\gcd(p_1, p_2, p_3) = \gcd(p_1, p_2)$, then

$$p_3 \leqslant f(p_1, p_2, p_3) \leqslant \max\{p_3, p_1 + p_2 - \gcd(p_1, p_2)\}.$$

In other words, if we have a word having periods p_1 and p_2, the supplementary hypothesis that w has also period p_3, can only improve the bound on the length of w given by the classical Fine and Wilf theorem (under the hypothesis that p_3 does not modify the gcd.)

Remark 4.3. The bound given in Theorem 4.2 is tight, as shown by the following example.
An infinite family of words proving the tightness will be studied in the next section.

Example 4.1. Consider the word

$$w = abacabaabacaba.$$

The word w has length $|w| = 14$ and periods $7, 11, 13$.

Since $\gcd(7, 11, 13) = h(7, 11, 13) = 1$, (cf. Example 3.1) one has

$$f(7, 11, 13) = \tfrac{1}{2}(7 + 11 + 13 - 2 + 1) = 15.$$

Then w is a word having periods $7, 11, 13$; its length is $|w| = f(7, 11, 13) - 1$ and w has no period $\gcd(7, 11, 13)$.

Remark 4.4. Let $I(n) = \{1, 2, \ldots, n\}$ be the set of positions of letters in a word $w = a_1 a_2 \ldots a_n$ of length n. By definition, w has period p if, for all $i, j \in I(n)$ such that $i \equiv j \pmod{p}$, $a_i = a_j$, i.e., at the positions i and j there is the same letter.

Let $[\text{mod } p_i]_n, i = 1,2,3$, denote the restriction to $I(n)$ of the equivalence mod p_i and let

$$R(p_1, p_2, p_3, n) = [\text{mod } p_1]_n \vee [\text{mod } p_2]_n \vee [\text{mod } p_3]_n.$$

Theorem 4.2 can be restated as follows:

If $n \geqslant f(p_1, p_2, p_3)$, then $R(p_1, p_2, p_3, n)$ coincides with $[\text{mod}(\gcd(p_1, p_2, p_3))]_n$, i.e., the restriction to $I(n)$ of the equivalence $\text{mod}(\gcd(p_1, p_2, p_3))$.

In particular, for $n \geqslant f(p_1, p_2, p_3)$, the number of classes of $R(p_1, p_2, p_3, n)$ is $\gcd(p_1, p_2, p_3)$. As a consequence the maximal number of different letters occurring in w is less than or equal to $\gcd(p_1, p_2, p_3)$.

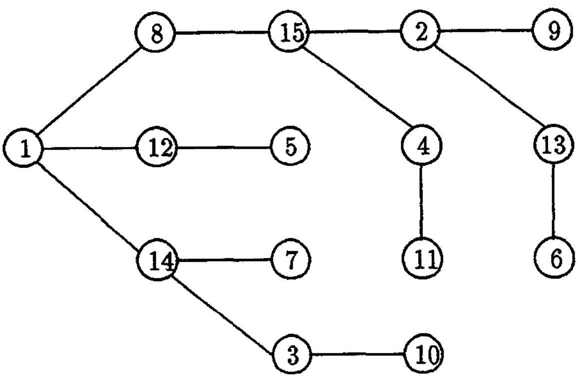
It will be useful in the sequel to translate this statement in terms of graphs. Let $G(p_1, p_2, p_3, n)$ be the graph having as set of nodes the set $I(n)$ and as edges the pairs (i,j) such that $|i - j| \in \{p_1, p_2, p_3\}$. If $n \geqslant f(p_1, p_2, p_3)$ the number of connected components of $G(p_1, p_2, p_3, n)$ is $\gcd(p_1, p_2, p_3)$. The nodes in the same connected component corresponds to positions in the word w carrying the same letter.

Note further that the node corresponding to the integer $f(p_1, p_2, p_3)$ has degree 3. Indeed, since $f(p_1, p_2, p_3) \geqslant \max\{p_1, p_2, p_3\}$, (cf. Remark 4.2), $f(p_1, p_2, p_3) - p_i$ belongs to $I(n)$, for $i = 1,2,3$.

Example 4.2. Let us consider the triple $(7, 11, 13)$ of previous examples.
The graph

$$G = G(7, 11, 13, f(7, 11, 13)) = G(7, 11, 13, 15)$$

is drawn in the following picture:



G is connected. The node labelled 15 has degree 3.

5. A generalization of Sturmian words

In this section we consider ordered triples (p_1, p_2, p_3) such that

$$\gcd(p_1, p_2, p_3) = h(p_1, p_2, p_3) = 1,$$

i.e., triples which are homogeneous and such that their elements are coprimes. Such triples are here called *good triples*.

Given a good triple (p_1, p_2, p_3), we have that

$$f(p_1, p_2, p_3) = \tfrac{1}{2}(p_1 + p_2 + p_3 - 1).$$

By Theorem 4.2, if a word w has periods p_1, p_2, p_3 and length

$$|w| \geqslant \tfrac{1}{2}(p_1 + p_2 + p_3 - 1),$$

then w is a power of a single letter, i.e., $w = a^n$, with $n = |w|$.

Let us define the set *3-PER* of the words w having periods p_1, p_2, p_3, where (p_1, p_2, p_3) is a good triple, and length

$$f(p_1, p_2, p_3) - 1 = \tfrac{1}{2}(p_1 + p_2 + p_3 - 3).$$

Remark 5.1. We claim that, for each good triple (p_1, p_2, p_3), the set *3-PER* contains exactly one word (up to a renaming), which moreover is ternary. Such a word can be constructed as follows. Let us consider the graph $G(p_1, p_2, p_3, f(p_1, p_2, p_3))$, introduced in Remark 4.4. As a consequence of Theorem 4.2, such a graph is connected. Let us consider, now, the graph $G' = G(p_1, p_2, p_3, f(p_1, p_2, p_3) - 1)$. G' is a subgraph of G, obtained by deleting the node corresponding to the integer $f(p_1, p_2, p_3)$. Since such a node has degree 3 (cf. Remark 4.4), the graph G' has at most three connected components. G' defines a word, of length $f(p_1, p_2, p_3) - 1$, in which the positions corresponding to the same connected component of G' carry the same letter. By construction, this word is the unique (up to a renaming) element of *3-PER* corresponding to the triple (p_1, p_2, p_3) and it is at most a ternary word.

Let us denote by $\mathscr{A} = \{a, b, c\}$ the alphabet of words in *3-PER*.

Example 5.1. The graph G', obtained from the graph G in Example 4.2 deleting the node 15, has three connected components:

$$C_1 = \{1, 3, 5, 7, 8, 10, 12, 14\},$$

$$C_2 = \{2, 6, 9, 13\},$$

$$C_3 = \{4, 11\}.$$

The graph G' defines the unique (up to a renaming) element of *3-PER* corresponding to the triple (7,11,13). It is the word

$$w = abacabaabacaba$$

of Example 4.1 and it is characterized as follows: the elements of C_1 corresponds to the positions of the letter a in the word w; the elements of C_2 corresponds to the positions of the letter b and those of C_3 to the positions of the letter c in the same word.

In [1] Arnoux and Rauzy studied, as a natural generalization of Sturmian sequences, the sequences over the alphabet $\{a,b,c\}$ having complexity $2n+1$ and satisfying the supplementary condition that, for any n, there exists only *one right special factor of degree* 3 of length n and only *one left special factor of degree* 3 of length n.

Denote by $L(AR)$ the language of the finite factors of the Arnoux and Rauzy sequences. The main result of this section is the following:

Theorem 5.1. $L(AR)=F(3\text{-}PER)$.

A fundamental tool in the proof of the Theorem 5.1 is a construction, introduced by Rauzy (cf. [1]).

Let us consider, for $i=1,2,3$, the maps $\varrho_i:(\mathscr{A}^*)^3 \to (\mathscr{A}^*)^3$, defined as follows:

$$\varrho_1(u,v,w)=(u,uv,uw),$$
$$\varrho_2(u,v,w)=(vu,v,vw),$$
$$\varrho_3(u,v,w)=(wu,wv,w).$$

Definition 5.1. The set $\mathscr{R}$ of the standard triples is the smallest subset of triples of words which satisfies the conditions:
(i) $(a,b,c)\in\mathscr{R}$;
(ii) $\mathscr{R}$ is closed under $\varrho_1,\varrho_2,\varrho_3$.

By the definition, any standard triple (u,v,w) can be expressed as

$$(u,v,w)=\varrho_{i_1}(\varrho_{i_2}(\dots\varrho_{i_k}(a,b,c)\dots)),\ \ i_1,\dots,i_k\in\{1,2,3\}.$$

The integer k is called the *depth* of the standard triple (u,v,w) and it is denoted by $d(u,v,w)$.

Remark 5.2. The reader can easily verify that, if (u,v,w) is a standard triple of words, then $(|u|,|v|,|w|)$ is a good triple of integers. Moreover, for any good triple (p_1,p_2,p_3) there exists a standard triple of words (u,v,w) such that $|u|=p_1$, $|v|=p_2$, $|w|=p_3$.

Lemma 5.1. *Let* $(u,v,w)\in\mathscr{R}$ *be a standard triple. The words in the set*

$$Perm(u,v,w)=\{uvw,uwv,vuw,vwu,wuv,wvu\}$$

have a common prefix of length $\frac{1}{2}(|u|+|v|+|w|-3)$.

Proof. The proof is obtained by induction on the depth $d(u,v,w)$ of the triple (u,v,w).

For $d(u,v,w)=1$, let us consider the triples $\rho_i(a,b,c)$, $i=1,2,3$.

Let us consider the case $i=1$:

$$\rho_1(a,b,c)=(a,ab,ac).$$

The words in the set $Perm(a,ab,ac)=\{aabac,aacab,abaac,abaca,acaab,acaba\}$ have as common prefix of length $\frac{1}{2}(2+2+1-3)=1$ the word a.

In a similar way, one verifies the cases $i = 2, 3$. Let us now suppose that the statement is true for the triple $(u, v, w) \in \mathcal{R}$ of depth n and consider, for instance, the triple $\rho_2(u, v, w) = (vu, v, vw)$. Similar arguments can be applied to the triples $\rho_1(u, v, w)$ and $\rho_3(u, v, w)$. Let us consider the set

$$Perm(vu, v, vw) = \{vuvvw, vuvwu, vvuvw, vvwvu, vwvuv, vwvvu\}.$$

We have to prove that such words have a common prefix of length

$$k = \tfrac{1}{2}(|vu| + |v| + |vw| - 3) = \tfrac{1}{2}(|u| + |v| + |w| - 3) + |v|.$$

The words in the set $Perm(vu, v, vw)$ have as a common prefix the word v. So it remain to prove that the words of the set

$$\{uvvw, uvwu, vuvw, vwvu, wvuv, wvvu\}$$

have a common prefix of length $h = k - |v| = \frac{1}{2}(|u| + |v| + |w| - 3)$. Let z be a word and m an integer. Let us denote by $pref_m(z)$ the prefix of length m of the word z. By convention, if $m \geqslant |z|$, $pref_m(z) = z$, and, if $m \leqslant 0$, $pref_m(z) = \varepsilon$. By using the inductive hypothesis, we have:

(i) $pref_h(uvwu) = pref_h(wvuv)$.

Moreover

$$\begin{aligned}
&\text{(ii)}\ pref_h(uvvw) = pref_h(uv)pref_{h-|uv|}(vw) = pref_h(uv)pref_{h-|uv|}(wv)\\
&\qquad = pref_h(uvwv),\\
&\text{(iii)}\ pref_h(vuvw) = pref_h(vu)pref_{h-|vu|}(vw) = pref_h(vu)pref_{h-|uv|}(wv)\\
&\qquad = pref_h(vuwv) = pref_h(uvwv),\\
&\text{(iv)}\ pref_h(vwvu) = pref_h(vw)pref_{h-|vw|}(vu) = pref_h(vw)pref_{h-|vw|}(uv)\\
&\qquad = pref_h(vwuv) = pref_h(uvwv),\\
&\text{(v)}\ pref_h(wvvu) = pref_h(wv)pref_{h-|wv|}(vu) = pref_h(wv)pref_{h-|wv|}(uv)\\
&\qquad = pref_h(wvuv) = pref_h(uvwv).
\end{aligned}$$

This conclude the proof. □

Definition 5.2. A word s is a 3-standard word if there exists a standard triple (u, v, w) such that s is the common prefix of the words in the set $Perm(u, v, w)$ of length $\frac{1}{2}(|uvw| - 3)$. Let us denote by 3-*ST* the set of 3-standard words.

Example 5.2. The triple

$$(u, v, w) = (abacaba, abacabaabacab, abacabaabac)$$

is a standard triple and it can be expressed as

$$(u, v, w) = \varrho_1(\varrho_3(\varrho_2(\varrho_1(a, b, c)))).$$

The depth of (u, v, w) is $d(u, v, w) = 4$. The common prefix of the words in the set $Perm(u, v, w)$ of length $\frac{1}{2}(|uvw| - 3) = 14$ is the standard word

$$s = abacabaabacaba.$$

This example also shows that the length of the common prefix in Lemma 5.1 is tight. In fact, one can verify that the words in the set $Perm(u,v,w)$ do not have a common prefix of length $\frac{1}{2}(|uvw|-3)+1=15$.

In order to prove the main result of this section, we need the following theorem, which is a consequence of a construction in [1] and which is here reported without proof.

Theorem 5.2. *F(3-ST)=L(AR).*

Next result states the equivalence between 3-standard words and the elements of the set *3-PER.*

Theorem 5.3. *3-ST = 3-PER.*

Proof. We first prove the inclusion *3-ST* $\subset$ *3-PER.*

Let s be a 3-standard word and let (u,v,w) be the standard triple such that s is the prefix of uvw of length $\frac{1}{2}(|uvw|-3)$. Let $|u|=p_1$, $|v|=p_2$, $|w|=p_3$. By Remark 5.2, (p_1,p_2,p_3) is a good triple. We prove that s has periods p_1, p_2, p_3, (and length $\frac{1}{2}(p_1+p_2+p_3-3)$), i.e., s belongs to *3-PER.* The proof is by induction on the depth n of the triple (u,v,w). For $d(u,v,w)=1$, we consider the triples $\rho_i(a,b,c)$, $(i=1,2,3)$. In the case $i=1$, $\rho_1(a,b,c)=(a,ab,ac)$. The corresponding 3-standard word s is the prefix of $aabac$ of length $\frac{1}{2}(5-3)=1$, i.e., $s=a$ and s has trivially periods 1 and 2. The cases $i=2,3$ are analogous.

Let us now consider a standard triple (u,v,w), with $|u|=p_1$, $|v|=p_2$, $|w|=p_3$, and let us suppose, by the inductive hypothesis, that the prefix s of uvw of length $\frac{1}{2}(p_1+p_2+p_3-3)$ has periods p_1, p_2, p_3. Take into account the standard triples $\rho_i(u,v,w)$, $i=1,2,3$. We consider the case $i=3$. The cases $i=1,2$ are analogous.

Let

$$\rho_3(u,v,w)=(wu,wv,w)=(u',v',w')$$

and let s' be the prefix of $u'v'w'=wuwvw$ of length $\frac{1}{2}(|u'v'w'|-3)$.
Since

$$\tfrac{1}{2}(|u'v'w'|-3)=\tfrac{1}{2}(p_1+p_2+3p_3-3)=\tfrac{1}{2}(p_1+p_2+p_3-3)+p_3,$$

one has that

$$|s'|=|s|+p_3.$$

We have to prove that s' has periods p_1+p_3, p_2+p_3 and p_3.

By definition

$$s'=pref_{|s|+p_3}(wuwvw)=w\cdot pref_{|s|}(uwvw).$$

By Lemma 5.1,

$$pref_{|s|}(uwvw)=pref_{|s|}(wuvw)=pref_{|s|}(uvww)=s.$$

Hence, $s' = ws$ and w is the prefix of s of length p_3. By Lemma 2.2, s' has periods $p_1 + p_3$, $p_2 + p_3$, p_3. This conclude the proof of the inclusion *3-ST* $\subset$ *3-PER*.

In order to state the converse inclusion *3-PER* $\subset$ *3-ST*, we observe that, for each good triple (p_1, p_2, p_3), there exists an element of *3-ST* having periods p_1, p_2, p_3 (cf. Remark 5.2). By Remark 5.1, for each good triple (p_1, p_2, p_3), *3-PER* contains exactly one word (up to a renaming). Since *3-ST* is trivially closed under the renaming of letters, we conclude that *3-PER* $\subset$ *3-ST*. $\square$

The proof of Theorem 5.1 is a consequence of Theorems 5.2 and 5.3.

Acknowledgements

We are very grateful to Sergio Salemi for helpful discussions and suggestions.

References

[1] P. Arnoux, G. Rauzy, Représentation géométrique de suites de complexité $2n + 1$, Bull. Soc. Math. France 119 (1991) 199–215.
[2] P. Arnoux, C. Mauduit, I. Shiokawa, J. Tamura, Complexity of sequences defined by billiard in the cube, Bull. Soc. Math. France 122 (1994) 1–12.
[3] J. Berstel, A. de Luca, Sturmian words, Lyndon words and trees, Theoret. Comput. Sci. 178 (1997) 171–203.
[4] J. Berstel, D. Perrin, Theory of Codes, Academic Press, New york, 1985.
[5] J.H. Bradley, Algorithm and bound for the greatest common divisor of n integers, Commun. ACM 13 (1970) 433–436.
[6] A. de Luca, F. Mignosi, Some combinatorial properties of Sturmian words, Theoret. Comput. Sci. 136 (1994) 361–385.
[7] A. de Luca, Sturmian words: structure, combinatorics and their arithmetics, Theoret. Comput. Sci. 183 (1997) 45–82.
[8] A. de Luca, Standard Sturmian morphisms, Theoret. Comput. Sci. 178 (1997) 205–224.
[9] N.J. Fine, H.S. Wilf, Uniqueness Theorem for Periodic Functions, Proc. Am. Math. Soc. (16) (1965) 109–114.
[10] R.L. Graham, D.E. Knuth, O. Patashnik, Concrete Mathematics, Addison-Wesley, Reading, MA, 1988.
[11] L.J. Guibas, A.M. Odlyzko, Periods in strings, J. Combin. Theory Ser. A (30) (1981) 19–42.
[12] G.H. Hardy, E.M. Wright, An Introduction to the Theory of Numbers, Fifth Edition, Clarendon Press, Oxford, 1938.
[13] P. Hubert, Complexité de suites définies par des billards rationnels, Bull. Soc. Math. France, 123 (1995) 257–270.
[14] P. Hubert, Propriétés combinatoires des suites définies par le billard dans les triangles pavants, Theoret. Comput. Sci. 164 (1996) 165–183.
[15] D.E. Knuth, The Art of Computer Programming, vol. 1, Addison-Wesley, Reading, MA, 1968.
[16] D.E. Knuth, The Art of Computer Programming, vol. 2, Addison-Wesley, Reading, MA, 1969.
[17] M. Lothaire. Combinatorics on Words, Addison-Wesley, Reading, MA, 1983.
[18] F. Mignosi, P. Séébold, Morphismes sturmiens et régles de Rauzy, J. théorie des nombres de Bordeaux 5 (1993) 221–233.
[19] G. Rauzy, Mots infinis en arithmétique, Lecture Notes in Computer Science, vol. 192, 1984, pp. 165–171.
[20] M.S. Waterman, Multidimensional greatest common divisor and Lehmer algorithms, BIT 17 (1977) 465–478.